

EXHIBIT A-1

In the Matter Of:

FREEMAN v

DEEBS-ELKENANEY

PATRICK JUOLA, PH.D.

September 26, 2023



1
2 UNITED STATES DISTRICT COURT
3 SOUTHERN DISTRICT OF NEW YORK
4 -----

5 LYNNE FREEMAN, an individual,

6 Plaintiff,

7 vs. Civil Action No.
8 1:22-cv-02435-LLS-SN

9 TRACY DEEBS-ELKENANEY P/K/A
10 TRACY WOLFF, an individual,
11 EMILY SYLVAN KIM, an individual,
12 PROSPECT AGENCY, LLC, a New Jersey
13 Limited Liability Company, ENTANGLED
14 PUBLISHING, LLC, a Delaware Limited
15 Liability Company, HOLTZBRINCK PUBLISHERS,
16 LLC D/B/A MACMILLAN, a New York Limited
17 Liability Company, and UNIVERSAL STUDIOS,
18 LLC, a Delaware Limited Liability Company,

19 Defendants.
20 -----

21
22 REMOTE VIDEOTAPED DEPOSITION OF
23 DR. PATRICK JUOLA, PH.D.

24 Tuesday, September 26, 2023

25 12:57 p.m. (EDT)

Reported By:

Joan Ferrara, RMR, FCRR

Job No. 2023-911534

September 26, 2023

12:57 a.m. (EDT)

Videotaped Deposition of DR. PATRICK
JUOLA, PH.D., held remotely via Zoom,
before Joan Ferrara, a Registered Merit
Reporter, Federal Certified Realtime
Reporter and Notary Public.

1
2 REMOTE APPEARANCES:

3
4 ON BEHALF OF PLAINTIFF:

5 DONIGER BURROUGHS

6 603 Rose Avenue

7 Venice, California 90291

8 BY: STEPHEN DONIGER, ESQ.

9
10
11 ON BEHALF OF PLAINTIFF:

12 CSREEDER

13 11766 Wilshire Boulevard

14 Suite 1470

15 Los Angeles, California 90025

16 BY: MARK PASSIN, ESQ.

17
18
19 ON BEHALF OF PROSPECT AGENCY and EMILY

20 SYLVAN KIM:

21 KLARIS LAW

22 29 Little West 12th Street

23 New York, New York 10014

24 BY: LANCE KOONCE, ESQ.

REMOTE APPEARANCES: (Continued)

ON BEHALF OF TRACY DEEBS-ELKENANEY P/K/A
TRACY WOLFF, ENTANGLED PUBLISHING, LLC,
HOLTZBRINCK PUBLISHERS, LLC d/b/a MACMILLAN
and UNIVERSAL CITY STUDIOS LLC:
COWAN DEBAETS ABRAHAMS & SHEPPARD

41 Madison Avenue

37th Floor

New York, New York 10010

BY: BENJAMIN HALPERIN, ESQ.

ALSO PRESENT:

Adrian Beltran, Videographer

Malcolm Coulthard

1 P. JUOLA

2 share substantial similarity, which would
3 have meant that if I studied several of
4 them, there would have been some things
5 that would have been overrepresented in the
6 statistics of the BMR versions, and as a
7 result would have violated the independent
8 assumptions that I was using in my report.

9 Q And one thing you say in your
10 report -- and I promise you, we will
11 introduce your report and talk about it
12 later, but one thing you say is you wanted
13 to avoid oversampling non-independent
14 texts.

15 Do you recall saying?

16 A I do recall saying that.

17 Q Why is it important to avoid
18 oversampling non-independent texts?

19 A For the reasons that I just
20 described. If you oversample things that
21 are not independent, then the properties of
22 the set from which these things are derived
23 will be either magnified in the case of
24 positive attributes or reduced in the case
25 of negative attributes.

1 P. JUOLA

2 And if you would like an analogy
3 for that, if I watch a wide receiver make a
4 marvelous touchdown catch on TV, he scores
5 6 points. He still only scores 6 points no
6 matter how often they show the instant
7 replay because the instant replay is not
8 independent of the actual footage of the
9 catch.

10 Q So does avoiding oversampling on
11 independent texts help ensure that your
12 data are not biased?

13 A Yes.

14 Q And is it important to ensure
15 that your data are not biased when you do a
16 statistical analysis?

17 A Yes.

18 Q Which specific books from the
19 Crave series written by Tracy Wolff are at
20 issue in this case?

21 A I don't know.

22 Q Is one of the books that you
23 considered titled "Crave"?

24 A Yes.

25 Q Is another entitled "Crush"?

P. JUOLA

A Yes.

Q Is another entitled "Covet"?

A Yes.

Q And is the last one entitled
"Court"?

A I don't remember.

Q I will represent to you that
those are the titles of the four Tracy
Wolff books at issue in this lawsuit.

A Okay. Thank you.

Q Did you read any of those four
Tracy Wolff books in developing your
opinions?

A I read them through my computer,
in the same sense that I read or did not
read BMR.

Q So that means you didn't plug
them on a Kindle and actually read them the
way you would read a book for pleasure?

A I did not read them in the same
way that I would read a book for pleasure.

Q You did not read them cover to
cover?

A I did not.

1 P. JUOLA

2 opining on substantial similarity. So
3 whether you want that answer to come
4 from him or you, Steve, I would
5 appreciate an answer now.

6 MR. DONIGER: I mean, I have the
7 same objection, is, you know, how we
8 end up arguing the relevance and
9 utility of his testimony is -- you
10 know, is to be determined.

11 His testimony is what it is. I
12 don't think we have a better answer
13 for you, Ben.

14 MR. HALPERIN: Okay. We can
15 move on. Thank you.

16 BY MR. HALPERIN:

17 Q Dr. Juola, you became famous
18 back in 2013 for uncovering that the book A
19 Cuckoo's Calling, which was published under
20 the name Robert Galbraith, spelled
21 G-A-L-B-R-A-I-T-H, was actually written by
22 Harry Potter author J.K. Rowling, correct?

23 A Yes, that is correct.

24 Q And I just want to be honest and
25 say I find this fascinating. So can you

1 P. JUOLA

2 please provide a play-by-play of how that
3 came to be and the steps you took in
4 figuring out that it was actually J.K.
5 Rowling who wrote that book?

6 A Well, if it would be helpful,
7 there is an article from Scientific
8 American in 2013 written by myself that
9 provides you with a very detailed
10 play-by-play, and it was written
11 contemporaneously, so it is less subject to
12 the passage of time and availability of
13 human memory.

14 In broad terms, the Sunday Times
15 of London telephoned me and said that
16 somebody had told them that J.K. Rowling
17 was actually the person behind The Cuckoo's
18 Calling and would I please investigate.

19 And so after discussion with Cal
20 Flynn, who was the reporter who tried to --
21 who had gotten in touch with me, we
22 obtained copies of The Casual Vacancy,
23 which was her first novel -- sorry, which
24 was her first adult novel -- The Cuckoo's
25 Calling, which was the questioned document

1 P. JUOLA

2 under investigation, and three other crime
3 novels by contemporary female British
4 authors.

5 One of them was a Ruth Rendell
6 novel, one of them was a P.D. James novel
7 and one of them was a Val McDermid novel.
8 And you can consult the Scientific American
9 article if you need to know exactly which
10 books they were.

11 I did a series of experiments
12 using a piece of software that I and my
13 team had developed to assess the similarity
14 in writing style between -- between The
15 Casual Vacancy -- sorry, excuse me --
16 between The Cuckoo's Calling and the four
17 books of known authorship.

18 And I found that all three of
19 the non-Rowling authors showed at least one
20 significant difference that disproved her
21 authorship -- that disproved their
22 authorship. So out of those four, the only
23 one that it could have been was Rowling.

24 Of course, there's the
25 possibility of none of the above, but I was

P. JUOLA

able to do some mathematics that established that the chance of it being none of the above was relatively low, and so I told -- I told the -- I told Ms. Flyn and the Sunday Times that it was probably Rowling.

The Times ran with the story, and Rowling admitted that she had written the -- she had written The Cuckoo's Calling.

Q Congratulations for that success.

A Thank you.

Q I'll introduce an exhibit now into the chat. Let me know when you see it and have it downloaded.

(Exhibit 2, article entitled, "How a Computer Program Helped Show J.K. Rowling write A Cuckoo's Calling" by Patrick Juola on August 12, 2013; 7 pages, was remotely introduced and provided electronically to the reporter, as of this date.)

A Yes, I have it downloaded.

1 P. JUOLA

2 Q We can call this one Juola
3 Exhibit 2.

4 Is this the Scientific American
5 article that you just referenced?

6 A It appears to be. If you made
7 some subtle changes -- if you made some
8 subtle changes to me, I probably wouldn't
9 detect them.

10 Q I will represent to you that we
11 have not made any changes to any of your
12 works that we'll discuss today.

13 A Thank you.

14 Q So this is an article published
15 in Scientific American entitled "How a
16 Computer Program Helped Show J.K. Rowling
17 Write The Cuckoo's Calling" -- shouldn't
18 that be "wrote," not "write"?

19 A I didn't write the headline, I'm
20 afraid.

21 If we're going to be -- if we're
22 going to be English teachers, The Cuckoo's
23 Calling should also be in italics.

24 Q And then it says: "By Patrick
25 Juola on August 20, 2013." Correct?

1 P. JUOLA

2 A Uh-huh.

3 Q So other than the --

4 A Yes.

5 Q Okay. So other than the
6 possible incorrect word and lack of
7 italics, this is an article that you wrote
8 on August 20, 2013, correct?

9 A The subhead is also not mine.

10 Q And this tells the story of how
11 you figured out it was J.K. Rowling who
12 wrote the book A Cuckoo's Calling?

13 A It does.

14 Q Please go to page 2.

15 A I am on page 2.

16 Q So I'm looking at the bottom
17 paragraph, and it says: "Over the past
18 decade, I have developed a computer program
19 to do this sort of analysis of writing
20 style, based on literally millions of
21 different features. This program will take
22 a sample of writing and determine, on the
23 basis of similarity, who among a set of
24 authors was most likely to have written
25 that sample."

1 P. JUOLA

2 Did I read that correctly?

3 A Yes, you did.

4 Q What's the name of this program?

5 A The name of this program is the
6 java authorship -- the Java Graphical
7 Authorship Attribution Program, abbreviated
8 J-G-A-A-P or JGAAP for short.

9 Q You mentioned this program in
10 your expert report, correct?

11 A Yes, I did.

12 Q Did you use J-G-A-A-P for your
13 analysis in this lawsuit?

14 A I did not.

15 Q You say in Scientific American
16 that J-J-A-P -- I'm sorry -- J-J-A-A-P --

17 A JGAAP.

18 Q Can we call it "JGAAP" and that
19 will be easier for me?

20 A Please.

21 Q I just messed it up twice, so --
22 anyways, strike all that and I'll rephrase.

23 You say in this paragraph that
24 JGAAP has literally millions of different
25 features that it can analyze, correct?

1 P. JUOLA

2 longer overlaps.

3 Q But this was not through using
4 your program JGAAP, correct?

5 A That is correct.

6 Q So at the end of the day, if you
7 were -- if this case were to go to trial
8 and you were to testify under oath on the
9 witness stand, would you tell the jury that
10 you used the same analysis that you did in
11 the Harry Potter case?

12 A I would not.

13 Q Can you please go back to page 3
14 of the Scientific American article, and let
15 me know when you're there.

16 A I'm there.

17 Q So I'm looking at the second
18 paragraph, and I'll read it. It says:
19 "Language use is a set of personal choices.
20 For example, the English language provides
21 a tremendous number of choices for words to
22 describe something bigger-than-big, words
23 such as 'huge,' 'giant,' 'enormous,' or
24 'colossal.' Writers can choose to express
25 an idea with a few precise words or a bunch

1 P. JUOLA

2 Q I'll just let you tell me when
3 you're ready.

4 A I'm ready.

5 Q Okay. Is this a publication
6 called "Authorship Attribution" by you,
7 Dr. Juola?

8 A It is.

9 Q Okay. And you called this -- or
10 you called something your 2006 book.

11 Is that what this is?

12 A That is what this is.

13 Q Okay. So I see two dates to the
14 top left of the front page. One says
15 Volume I, No. 3, 2006, and then copyright
16 2008.

17 But just for our purposes, we
18 can call this your 2006 book called
19 "Authorship Attribution."

20 A It took a while to come out,
21 yes.

22 Q Sure.

23 And just to read from the
24 abstract, it says (as read): Authorship
25 attribution, the science of inferring

1 P. JUOLA

2 characteristics of the author from the
3 characteristics of documents written by
4 that author, is a problem with a long
5 history and a wide range of application.
6 Recent work in non-traditional authorship
7 attribution demonstrates the practicality
8 of automatically analyzing documents based
9 on authorial style, but the state of the
10 art is confusing. Analyses are difficult
11 to apply, little is known about the type or
12 rate of errors, and few best practices are
13 available. In part because of this
14 confusion, the field has perhaps had less
15 uptake and general acceptance than is its
16 due.

17 Did I get through that okay?

18 A Yes.

19 Q And so one of the things you
20 wrote there is that little is known about
21 the type or rate of errors for authorship
22 attribution, correct?

23 A Yes.

24 Q Do you ever recall writing that
25 stylometry is much less reliable and

1 P. JUOLA

2 accurate than DNA analysis?

3 A I remember writing similar
4 statements. I don't remember that one
5 specifically.

6 Q Why don't we go back to the
7 Scientific American article, which I
8 believe is Exhibit 2.

9 A Got it.

10 Q And go to page 6.

11 A Yes. I am there.

12 Q Okay. And so I just wanted to
13 verify that the phrase at the top of
14 page 6, "Stylometry is much less reliable
15 and accurate than DNA," that's something
16 you wrote in this Scientific American
17 piece, correct?

18 A Yes.

19 Q So do you agree that there are
20 not clearcut standards for applying
21 stylometry to determine whether two works
22 share common authorship?

23 A Yes, I did.

24 Q Do you recall ever writing that
25 new best practices are available for

1 P. JUOLA

2 stylometry?

3 A Again, I've written that -- I've
4 written statements like that. I don't
5 remember that specific wording.

6 Q That was actually part of the
7 statement from your 2006 book that we just
8 read.

9 A Okay. If it's in my book, then
10 I wrote it.

11 Q Do you ever -- do you recall
12 ever writing that one of the major
13 weaknesses in the application of authorship
14 attribution was the absence of clearcut
15 standards for accurate analytical practice?

16 A Again, I have written things
17 like that. I don't remember that specific
18 wording.

19 Q All right. I'll just show it to
20 you because I'm not trying to quiz your
21 memory. I just want to verify it.

22 So let's introduce one more of
23 your writings, and this is called the
24 Rowling Case --

25 A Yes.

P. JUOLA

Q -- 2015.

Let me know when you have it open.

A Okay.

MR. HALPERIN: We're going to mark this Juola Exhibit 7.

(Exhibit 7, article entitled, "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions," by Patrick Juola; 14 pages, was remotely introduced and provided electronically to the reporter, as of this date.)

BY MR. HALPERIN:

Q Are you looking at the abstract?

A I am now.

Q So it says there: "We propose" -- and I'm just reading from the abstract -- "We propose a possible solution to one of the major weaknesses in the application of authorship attribution - the absence of clearcut standards for accurate analytic practice."

Did I read that correct?

1 P. JUOLA

2 A Yes.

3 Q And that's something you wrote
4 in this publication from 2015?

5 A Yes.

6 Q How many words long is the
7 manuscript of BMR that you considered?

8 A I would have to refresh my
9 memory on that. It's novel length.

10 Q I believe you say in your report
11 that it's over 150,000 words.

12 Does that sound right to you?

13 A That sounds about right.

14 Q Do you recall collectively how
15 long each of the four creative books at
16 issue in this case are?

17 A Not -- not specifically. Again,
18 they're novels.

19 Q In paragraph 27 of your report,
20 you say they collectively have 868,455
21 words.

22 A Okay.

23 Q Does that sound right?

24 A That sounds about right.

25 Q Sir, would you agree that

1 P. JUOLA

2 A Yes.

3 Q When you have one seven --
4 common seven-word sequence that is shared
5 in two novel-length works, can you put a
6 percentage on how likely you are to be
7 right as to whether there is common
8 authorship?

9 MR. DONIGER: Objection.

10 Incomplete hypothetical.

11 A It would take a lot of math. It
12 might be possible to do.

13 I did not do it for this report.

14 Q Let's look at paragraph 19 of
15 your report, please.

16 A Yes.

17 Q You're there?

18 A I'm at 19.

19 Q All right. I'll just read this
20 paragraph in full first.

21 "We did find several such
22 phrases shared between Crave and BMR, of
23 which 'in the air as I try to,' 'that
24 million-dollar smile of his,' and 'on my
25 arms and the back of my neck' are the most

1 P. JUOLA

2 notable. (Others, such as 'the ground to
3 open up and swallow me' can be easily
4 explained as a common idiom/expression.)
5 The phrase 'on my arms and the back of my
6 neck' appears in roughly 20 independent
7 contexts in Google search, which is rare
8 but still present. Similarly, 'that
9 million-dollar smile of his' is only six
10 words long but occurs only 3,370 times
11 among the hundreds of billions of web pages
12 indexed by the Google search engine. The
13 phrase 'in the air as I try to' appears
14 only seven times in this set."

15 Did I get through that
16 correctly?

17 A Yes, you did.

18 Q All right. So for the part of
19 your methodology that involved looking at
20 seven-word sequences, those are the three
21 phrases that you considered that
22 potentially qualified?

23 A Yes.

24 Q Did you identify those three
25 phrases yourself or did someone else?

1 P. JUOLA

2 A I identified those phrases
3 myself.

4 Q And if I understand correctly,
5 these weren't the only shared phrases you
6 found, but these were the ones that you
7 deemed most notable, correct?

8 A That is correct.

9 Q Okay. But the only -- for the
10 part -- and only talking about here the
11 part of your methodology that involved
12 seven-word sequences, the only phrases
13 identified in your report are the three we
14 just discussed and the one "the ground to
15 open up and swallow me," which it seems
16 like you rejected in your analysis?

17 A That is correct.

18 Q Sitting here today, are you able
19 to remember what the other phrases were?

20 A I do not.

21 Q And you're not opining that
22 shared phrases other than those specific
23 three listed in this paragraph are evidence
24 of the shared authorship, correct?

25 A I am not opining anything about

1 P. JUOLA

2 other shared phrases.

3 Q Let's talk about the phrases.

4 So the first one is "in the air
5 as I try to," correct?

6 A Correct.

7 Q And that one is seven words
8 long?

9 A Yes.

10 Q The next one is "that
11 million-dollar smile of his," correct?

12 A Correct.

13 Q How many words is that one?

14 A Depends on who you ask and how
15 you treat the hyphen.

16 Q So the hyphen might -- some
17 people might treat the hyphen as making it
18 one word, some people might treat the
19 hyphen as it being two words?

20 A That is correct.

21 My understanding is that the
22 Google search engine that I used treats the
23 hyphen as a word separator, and so that is
24 six words in the opinion of Google.

25 Q Okay.

1 P. JUOLA

2 So when a linguist looks at a
3 phrase like this, one thing they might do
4 is they might count "million-dollar" as two
5 separate words and call this a six-word
6 phrase. But would it also be reasonable to
7 count it as a five-word phrase?

8 A Yes.

9 Q Do you have a view as to which
10 camp is right and which camp is wrong about
11 that?

12 A Not really.

13 Q What's an idiom?

14 A An idiom is a phrase that is --
15 technically, an idiom is a phrase where the
16 meaning of the phrase as a whole does not
17 correspond to the meaning of the individual
18 words in it, such as if I say my
19 grandmother kicked the bucket in 19 -- in
20 2001, she wasn't actually kicking any
21 buckets.

22 Q Sure. So by that definition,
23 would "million-dollar smile" an idiom?

24 A I think "million-dollar smile"
25 is a metaphor and not an idiom. I think

1 P. JUOLA

2 A -- much larger than web pages.

3 Q So does this change your view at
4 all as to whether "million-dollar smile" is
5 a common phrase?

6 A It's starting to look like --
7 it's starting to look like something I
8 would call common.

9 Q If "million-dollar smile" is a
10 common phrase, what does that do to the
11 length of the phrase "that million-dollar
12 smile" hits?

13 A Then it would basically drop
14 down to the equivalent of a four-word
15 phrase --

16 Q Okay.

17 A -- treating "million-dollar
18 smile" as a single word.

19 Q Is "in the air" a common phrase?

20 A Again, I would want to look at
21 that.

22 Q Okay.

23 A Go ahead.

24 Q I'll do it right now. I have no
25 problem.

1 P. JUOLA

2 A Okay.

3 Q I'm putting it in quotes right
4 now. We're on the "Books" tab. We're only
5 going to see the "Books" tab first. Okay?

6 A Yeah, the "Books" tab is fine.

7 Q About 97,300,000 results.

8 A Okay. So that is in the text
9 because that's more than once per book.

10 Q Would you consider it a common
11 phrase based on this?

12 A Yeah, I'll give that to you.

13 Q I feel like I don't need to ask
14 this anymore, but I was so excited about
15 this question so I'm just going to do it
16 anyway.

17 Have you heard of the Phil
18 Collins song "In The Air Tonight"?

19 A Yes, I have.

20 Q Do you know what year that came
21 out?

22 A A long time ago.

23 Q 1981. Does that sound right?

24 A Yeah.

25 Q It came out a long time before

1 P. JUOLA

2 Blue Moon Rising or the Crave, right?

3 A Yes.

4 Q All right. So if "in the air"

5 is a common phrase, what does that do to

6 the length of "in the air as I try to"?

7 A That would reduce it from seven

8 words to five, effectively.

9 Q And your last example from

10 paragraph 19 is "on my arms and the back of

11 my neck," correct?

12 A Right.

13 Q And how many words is that?

14 A What?

15 Q How many words is "on my arms

16 and the back of my neck"?

17 A Is that 10? I mean, I don't

18 have it in the front of me to count.

19 Q I misheard you. Two, three,

20 four, five, six, seven, eight, nine.

21 A Nine. My apologies.

22 Q Okay. Would you like to see --

23 can we -- well, first, let me ask you this,

24 is "on my arms" a common phrase?

25 A "On my arms" is not that common,

1 P. JUOLA
2 but "the back of my neck" -- but "the back
3 of my neck" is. I will -- you will
4 probably be asking that question.

5 Q All right. So you would agree
6 that "back of my neck" is common?

7 A Yes.

8 Q Okay. So without making a
9 determination either way of "on my arms"
10 and "the back of my neck," if we consider
11 the four words "back of my neck" to be
12 common, what does that do to the sequence
13 here?

14 A Then it drops down to six words,
15 I believe.

16 Q Okay. And could we expand the
17 common sequence to "the back of my neck"?

18 A Maybe.

19 Q And if people are going to say,
20 "back of my neck," they're probably going
21 to put the word "the" --

22 A Yes, "the" would be the most
23 likely way to put it.

24 Q And it drops this one down to
25 five words, right?

1 P. JUOLA

2 A Yeah.

3 Q At the end of paragraph 19, you
4 write: "The phrase 'in the air as I try
5 to' appears only seven times in this set."

6 Do you see that?

7 A Uh-huh, yeah.

8 Q What set are you referring to?

9 A The hundreds of billions of web
10 pages indexed by the Google search engine.

11 Q Okay. I'm just going to go back
12 to Google.com again.

13 Do you see it on my screen?

14 A Uh-huh.

15 Q All right. So I'm just going to
16 search for that whole phrase now, "in the
17 air as I try to".

18 Did I do that all right?

19 A Yes.

20 Q I have it in quotes?

21 A Yes.

22 Q All right. So --

23 A Oh, they must have update their
24 database.

25 Q Okay. I'm seeing -- is it

1 P. JUOLA

2 possible by seven times, you meant 207,000
3 times?

4 A No.

5 Q So seven times in this set, just
6 off by a factor of over 200,000?

7 A 200,000, yeah.

8 Google -- Google has a tendency
9 to sometimes make major updates to their
10 database.

11 Q And so you think it was a major
12 update between May 10th and today that
13 caused the number to increase?

14 A Quite possibly.

15 Q And can you give us an equation
16 of how -- mathematical equation of how big
17 this difference is here?

18 A Well, it's an increase of
19 roughly 50,000 times.

20 Q Okay.

21 Let's go to paragraph 20. You
22 say: Thus, "in the air as I try to"
23 qualifies as a seven-word lexical overlap
24 and is, in Coulthard's view, unlikely to be
25 independently written by two different

1 P. JUOLA

2 A Oh, no. It looks like I deleted
3 an entire phrase. Even if they do not meet
4 Coulthard's seven-word criterion.

5 Q Okay. So you agree that the
6 five and six phrases do not meet
7 Dr. Coulthard's --

8 A No, they do not.

9 Q And if we were to knock all
10 these phrases down because we just
11 discovered that each one contains at least
12 one common phrase, then none of them meet
13 Dr. Coulthard's criteria, right?

14 A Well, Dr. Coulthard's criterion
15 does not specify how lexical bundles are to
16 be handled.

17 What I said about just reducing
18 them down to a single word is my rule of
19 thumb and not Dr. Coulthard's.

20 Q Okay.

21 A So I cannot -- I cannot answer
22 for Dr. Coulthard, and I believe he's your
23 witness anyway.

24 Q But under your criteria, you
25 would count them as less than seven

1 P. JUOLA

2 words --

3 A Yeah, under my criteria, I'd
4 count them as less than seven.

5 Q Well, let's talk about
6 Dr. Coulthard.

7 Have you worked with him before?

8 A Yes, I have.

9 Q Would you say he's
10 well-respected in forensic linguistics?

11 A I would.

12 Q Would you say he's one of the
13 most important scholars in forensic
14 linguistics?

15 A I would.

16 Q We discussed the International
17 Association of Forensic Linguists, which
18 now has another L.

19 A Uh-huh.

20 Q But you remember the
21 organization we were talking about?

22 A I do.

23 Q Okay. Are you aware that
24 Dr. Coulthard helped found the
25 International Association of Forensic

1 P. JUOLA

2 Since the only time I've spoken
3 with him is in the presence of the lawyers,
4 I think anything else is privileged.

5 Q Do you know whether he's Lynne
6 Freeman's husband?

7 A I believe that's privileged
8 information.

9 MR. HALPERIN: Steve, is that
10 privileged information?

11 MR. DONIGER: Dr. Juola, they
12 did depose Mr. Baer and are aware he's
13 Lynne Freeman's husband.

14 A It has been represented to me
15 that he's Lynne Freeman's husband. I have
16 not seen their marriage license.

17 Q He has testified under oath that
18 he is Lynne Freeman's husband. Is that
19 good enough for you?

20 A Since I don't actually care.

21 Q Okay.

22 Would you agree Mr. Baer has an
23 interest in whether or not the plaintiff
24 prevails in this lawsuit?

25 A I would indeed.

1 P. JUOLA

2 Q So you say that he sends us a
3 list of twenty-five commonalities that he
4 considered noteworthy, of which six related
5 specifically to the 2011 version of BMR
6 that we analyzed.

7 Did I read that right?

8 A Yes.

9 Q Okay. Do you know what
10 methodology Mr. Baer used to identify the
11 25 commonalities he sent you?

12 A I do not.

13 Q And do you know what methodology
14 Mr. Baer used to make a determination as to
15 whether or not those commonalities were
16 noteworthy?

17 A I do not.

18 Q Do you know whether Mr. Baer
19 took an evenhanded approach and looked for
20 both similarities and differences?

21 A I do not.

22 Q As a scientist, do you agree
23 that cherrypicking data is a bad thing?

24 MR. DONIGER: Calls for
25 speculation. Vague and ambiguous as

1 P. JUOLA

2 purposefully selecting data that favors
3 your desired result? Is that fair?

4 A Was that a question to Stephen
5 or to me?

6 Q It was to you. I was going to
7 ask if you would agree to that as a
8 definition for "cherrypicking."

9 A Can you repeat it, please?

10 Q Purposefully selecting data that
11 favors your desired result.

12 A I can work with that.

13 Q And you don't have any idea of
14 whether or not Mr. Baer engaged in
15 cherrypicking as we just defined it?

16 A I do not.

17 Q Do you know what Mr. Bear's
18 occupation is?

19 A I think he mentioned it once,
20 but I don't remember.

21 Q If I told you he testified under
22 oath that he is a financial advisor, would
23 you have any reason to dispute that?

24 A I would not have any reason to
25 dispute that.

1 P. JUOLA

2 books; and secondly, whether or not they
3 appeared in the Google Books Ngram
4 Database." Correct?

5 A Correct.

6 Q All right. So if I understand
7 your methodology here, it's kind of three
8 steps. You take one of the phrases, you
9 first search BMR for it. That's step one.

10 A Uh-huh.

11 Q Step two, you search one of the
12 four Crave books for it.

13 And then step three, you search
14 to see whether it's in the Google Books
15 Ngrams Database, correct?

16 A That is correct.

17 Q All right. Three-step
18 methodology that we can agree on?

19 A With one minor point, I did not
20 search just one of the Crave books.

21 Q I agree. I said all four.

22 And you searched any of the four
23 Crave books?

24 A I searched all four to see if it
25 occurred in any of them.

1 P. JUOLA

2 any sequence of five or fewer words and see
3 how often it appears by year.

4 Q Do you know who created the
5 Google Books Ngram's corpus?

6 A It was a huge team. The nominal
7 author is Jean-Baptiste. Jean-Baptiste
8 Camps, I think.

9 Q Jean-Baptiste Michel,
10 M-I-C-H-E-L?

11 A Oh, yes.

12 Q What is Google Books Ngrams
13 typically used for?

14 A Can't say. I'm not typical.

15 MR. DONIGER: Object as calling
16 for speculation. Lacking foundation.
17 BY MR. HALPERIN:

18 Q Would you say one of the core
19 uses of Google Books Ngrams is to
20 investigate language trends over time?

21 MR. DONIGER: Calls for
22 speculation.

23 A That is certainly a common use.

24 Q Do you think that is the most
25 common use of it, to analyze language

1 P. JUOLA

2 A I think it was published very,
3 very recently, 2018 or 2022. But the
4 project itself was in -- the project itself
5 was in publishing limbo for a very long
6 time.

7 Q If you were to go back and --
8 and you can check me if you want, but I'll
9 tell you -- I'll represent it. It says in
10 this at like the second page of the
11 encyclopedia that it was published in 2022.

12 A Yeah. I think that's what I've
13 got on my CV as well, so...

14 Q Okay. So on page 517, do you
15 recognize this as a chapter you wrote about
16 Google Books Ngrams?

17 A Yes.

18 Q All right. So you are the
19 Patrick Juola that it lists there --

20 A I am.

21 Q -- on 517 -- okay.

22 So I'm going to go to the next
23 page 518.

24 A Uh-huh, I'm there.

25 Q Okay. Great.

1 P. JUOLA

2 I'm looking at the bottom left
3 paragraph.

4 A Uh-huh.

5 Q I'm going to read it to you.

6 "In part to satisfy the demands
7 of copyright law, Google does not typically
8 make the full text of books available, but
9 is allowed to publish and distribute
10 snippets. The Google Books Ngram corpus
11 (Michel et al. 2011) provides n-grams
12 (groups of n consecutive non-blank
13 characters separated by whitespace) for
14 5 million books at values of n from 1 to
15 5."

16 Did I read that correctly?

17 A Yes, you did.

18 Q All right. So here you said it
19 contains 5 million books.

20 A Yeah, I believe that is the
21 number in 2011.

22 Q This isn't 2011. This is 2022.

23 A But the citation for Michel is
24 2011.

25 Q Oh. So in this chapter written

1 P. JUOLA

2 in this 2022 book, you said 5 million, but
3 you were basing that on data from 2011?

4 A Yes.

5 Q Okay. And so now you're saying
6 15 million in your report without a
7 citation, and that's because you believe
8 it's increased --

9 A Yes.

10 Q Was it three-fold since then?

11 A That would be three-fold.

12 Q Okay. Can you provide me today
13 with a citation for where that
14 15 million --

15 A Today, I cannot.

16 Q Okay. If the number were
17 5 million, not 15 million, that would
18 affect your calculations, correct?

19 A That would affect my
20 calculations.

21 Q All right.

22 Okay. And I believe you said
23 that in Google Books Ngrams, it contains
24 books, but it doesn't contain the entirety
25 of the books that the corpus comes from,

1 P. JUOLA

2 correct?

3 A I didn't say that.

4 Q Well, it only contains n-grams
5 that are up to five words long, right?

6 A Yes.

7 Q Okay. So it doesn't contain the
8 rest of the books, just only the n-grams
9 that are up to five words long?

10 A I don't understand the question.

11 Q Is every word from books -- from
12 a book that is included in Google Ngrams
13 included in the corpus?

14 A Every individual word is
15 included as a 1-gram.

16 Q Okay.

17 A Every pair of adjacent words is
18 included as a 2-gram.

19 Q Okay. Google Books Ngrams
20 includes books in multiple different
21 languages, correct?

22 A Yes.

23 Q It includes books in Chinese,
24 right?

25 A Yes.

1 P. JUOLA

2 Q French?

3 A Yes.

4 Q German, correct?

5 A Yes.

6 Q Italian, correct?

7 A I'm not -- I'm not sure. I

8 believe so. I'd have to re-check.

9 Q Do you know if Spanish is
10 included?

11 A Spanish is included.

12 Q Russian?

13 A Again, I'd have to check. I've
14 never used the Russian.

15 Q Hebrew?

16 A Again, I'd have to check.

17 Q Do you know how many books in
18 the Google Books Ngrams database are in
19 English?

20 A The majority of them.

21 Q Do you know what percentage in
22 the Google Books database are -- Ngrams
23 database are in English?

24 A Not offhand.

25 Q What is an n-gram?

1 P. JUOLA

2 MR. HALPERIN: And for Joan,
3 please let me clarify the spelling, N
4 dash G-R-A-M.

5 Q And Dr. Juola, you can answer.

6 A An n-gram, generally speaking,
7 is a sequence of n consecutive items --
8 sorry, it is a subsequence of n consecutive
9 items from a -- typically from a longer
10 sequence.

11 Q And it's correct that it's not
12 necessarily a sequence of words because it
13 could be, for example, like the word -- it
14 could be like "scuba" is one of them or
15 something, or some acronym, right?

16 Maybe that's confusing.

17 A Any linguist would consider
18 "scuba" a word. But I can also create
19 n-grams of parts of speech.

20 Q Okay.

21 A I could say, for example, the
22 cat sat on the mat, and that would be --
23 and if you looked just at the parts of
24 speech of that, that would be article,
25 noun, verb, preposition, article, noun.

1 P. JUOLA

2 Q Okay.

3 A And I can say that article noun
4 is a very common 2-gram in English. Noun
5 article is not. It's like saying "cat
6 the."

7 Q Just so we can understand
8 ourselves when we're talking about it
9 today, can we refer to an n-gram as a
10 consecutive sequence of words?

11 A Can't we refer to it as an
12 n-gram? That's the reason the term was
13 invented.

14 Q Okay. I just wanted to --
15 non-linguists are going to be reading your
16 transcript, so I just want to make sure
17 they understand it.

18 A Okay. An n-gram is a
19 consecutive sequence of words, unless
20 otherwise noted. If I need talk about an
21 n-gram of something else, I will make it
22 clear.

23 Q I appreciate that.

24 And Google Books Ngrams does not
25 include n-grams longer than five, correct?

1 P. JUOLA

2 A That is correct.

3 Q So any sequence of words longer
4 than five words will not be in the database
5 by definition, right?

6 A It will be -- any sequence of
7 six words, it will be in the database as
8 two partial sequences, words one through
9 five and words two through six.

10 Q But if you searched for the
11 entire six-word sequence, it would return
12 nothing, right?

13 A It would return nothing.

14 Q And so you actually couldn't use
15 Google Books Ngrams if you were trying to
16 follow what you described as
17 Dr. Coulthard's methodology because Google
18 Ngrams will not return any result for a
19 seven-word sequence, correct?

20 A That is correct.

21 Q So that would actually make
22 Google Books Ngrams a poor database to use
23 to search for common authorship if you were
24 trying to follow Dr. Coulthard's
25 methodology, as you described it, right?